2010 13th International IEEE
Annual Conference on Intelligent Transportation Systems
Madeira Island, Portugal, September 19-22, 2010

WA1.5

Short-Term Traffic Speed Forecasting Based On Data Recorded at Irregular Intervals

Qing Ye, S.C. Wong, W.Y. Szeto

Department of Civil Engineering, The University of Hong Kong

Abstract - **As demand for proactive real-time transportation management systems has grown, major developments have been seen in short-time traffic forecasting methods. Recent studies have introduced time series theory, neural networks, genetic algorithms, etc., to short-time traffic forecasting to make forecasts more reliable, efficient and accurate. However, most of these methods can only deal with data recorded at regular time intervals, thereby restricting the range of data collection tools to loop detectors or other equipment that generate regular data. The study reported here represents an attempt to expand on several existing time series forecasting methods to accommodate data recorded at irregular time intervals, thus ensuring these methods can be used to obtain predicted traffic speeds through intermittent data sources such as the GPS. The study tested several methods using the GPS data from 480 Hong Kong taxis. The results show that the best performance is obtained using a neural network model with acceleration information predicted by ARIMA model.**

I. INTRODUCTION

Irregularly spaced or intermittent data is a problem commonly encountered in many research fields. Many researchers have already extended the classic time series methods in an attempt to deal with this problem. Croston [1] developed a method for providing a more accurate estimate of stock control in 1972. This method separately applies exponential smoothing to the intervals between non-zero demands and the magnitude of the non-zero demands by treating them as a continuous time series according to their index. Jones [2] tried to use the maximum likelihood fitting of autoregressive moving average (ARMA) models for time series with missing observations in 1980. However, this method is useful only when the number of missing observations is very small. In 1986, Wright [3] further developed the simple exponential smoothing method and Holt's method by extending the use of weighted averages for irregularly spaced data. In 1995, Cipra et al. [4] expanded on Wright's modification of Holt's method by proposing an extended version of the Holt-Winters method for data taken at irregular time intervals. In 2004, Carmo and Rodrigues [5] compared the methods of Wright and Croston to a neural network method to show how the neural network can be used for filtering and forecasts. Willemain et al. [6] suggested a bootstrap method for intermittent data on service parts inventories in 2004. In the following year, Velicer and Colby [7] compared four different data-missing methods (the deletion, mean substitution, mean of adjacent observations

and maximum likelihood estimation methods) commonly used for autoregressive integrated moving average (ARIMA) models time series under 50 different data-missing conditions, their results showing that the maximum likelihood estimation method performed the best under all conditions tested. In 2006, Cipra and Praha [8] made a modification to the double smoothing method that was similar to that Wright made to Holt's method and suggested that the modified double smoothing method performed better than the modified Holt method. However, this comparison may not be fair, because Cipra and Praha did not optimize the coefficients in these methods. Altay et al. [9] compared a modification of Croston's method and Wright's modified version of Holt's method in 2008, suggesting that using Croston's method to forecast demand will result in the minimization of inventory levels, whereas using the modified Holt method will result in superior customer service. Cipra and Hanzak [10] derived a method involving the exponential smoothing of order m for irregular time series in 2008.

Many researchers have also used time series to forecast traffic flow in the transportation field. In 1997, Smith et al. [11] compared four methods – the time-series method and the historical average, neural network and nonparametric regression models – to estimate traffic flow 15 minutes ahead. The results indicated that the nonparametric regression model was the most accurate. However, in 2002 Smith et al. [12] reached a different conclusion in showing that the seasonal autoregressive integrated moving average (SARIMA) model delivers results that are statistically superior to those of the nonparametric models. William et al. [13] proposed the SARIMA model as a theoretical basis for modeling traffic data. Szeto et al. [14] developed an integrated method incorporating both the SARIMA model and the cell transmission model. The forecasting results suggested that these methods are effective in predicting real-time traffic flow data. Tan et al. [15] suggested an aggregation method in which the moving average (MA) and exponential smoothing (ES) methods and the ARIMA model are used to forecast three relevant time series and the predicted results are aggregated in a neural network model to output the final forecast.

The studies outlined above have together resulted in significant progress in forecasting traffic flow by treating traffic condition data as time series. However, most existing methods can only deal with regularly spaced time series data

because they involve the use of loop detectors to obtain the data required. Given that it is not possible to have loop detectors installed in all links in a city, some alternative ways of providing information on traffic conditions have been developed, such as floating cars which upload GPS data. The study reported here was an attempt to address the problem whereby using GPS data may result in irregular time intervals in a time series. It sought to expand on the classic ways of dealing with intermittent time series to allow for traffic conditions to be forecast using irregularly spaced data.

II. DATA

The study used the GPS data of 480 Hong Kong taxis collected over the 14 weeks from 29 June 2009 to 4 October 2009. The GPS equipment installed in the taxis uploaded their coordinates, speed, driving direction and the uploading time to the information centre approximately once every 30 seconds. The study combined the use of a GIS technique to locate the taxis' position according to their coordinates and on a road map. A 2-kilometer long section of Princess Margaret Road approaching the Cross Harbour Tunnel in Hong Kong was selected and separated into 20 100-meter long segments that were used as the 20 sample points. Data for the first 12 weeks were used for calibration and data for the last two weeks were used for validation. The data for the first 12 weeks comprised 181,952 observations. The maximum speed recorded was 120 kilometers per hour and the minimum speed recorded was 0 kilometers per hour, with a mean speed of 22.85 kilometers per hour and a standard deviation of 25.38. The data recorded in the last two weeks included 33,509 observations. The maximum speed recorded was 95 kilometers per hour and the minimum speed recorded was 0 kilometers per hour, with a mean speed of 20.58 kilometers per hour and a standard deviation of 24.26.

III. METHODS

The study intends to modify several existing methods to forecast speeds using irregularly spaced data. They are the Naïve method, Wright's modification of the simple exponential smoothing method, Wright's modification of Holt's method[3]. An extrapolation method using the latest observation and predicted acceleration was also used for comparison. Two neural network methods, neural network with and without acceleration information were also proposed to aggregate the result of the three existing methods. ARIMA model was used to predict the acceleration with a transformation to deal with the irregularity. The first three methods predicted the forecasts using speed data only; the following two methods used to import and combined the acceleration information.

Existing Parametric Methods

A. The Naïve Method

Under the Naïve method, the latest observation is used as the forecast for the next period. It serves as the worst-case scenario and can be considered a reference method to which more complex methods can be compared. It can be mathematically expressed as follows:

$$\hat{y}_t = y_t \tag{1}$$

where y_t is the speed observed at time t and $\hat{y}_t$ is the forecast made at time t.

B. The Simple Exponential Smoothing Method with Wright's Modification

This method represents a compromise between two extreme prediction methods: the Naïve method, in which all the weight is placed on the latest observation, and the moving average method, in which all the observations are weighted equally to obtain the average. The exponential smoothing method gives the most weight to the latest observation and gives exponentially decreasing weight to distant observations. All exponential smoothing methods follow a basic principle of recursive form:

$$new\ estimate = (1 - parameter) \times previous\ estimate + parameter \times new\ observation \tag{2}$$

The simple exponential smoothing method is the basic method that follows this recursive form.

The mathematic expression of the simple exponential smoothing method is:

$$\hat{y}_t = \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \alpha(1-\alpha)^3 y_{t-3} + ... \tag{3}$$

where y_t is the speed observed at time t, $\hat{y}_t$ is the forecast made at time t, and α is a coefficient ranging from 0-1.

The recursive form of the same method is:

$$\hat{y}_t = (1-\alpha)\hat{y}_{t-1} + \alpha y_t \tag{4}$$

However, the form given above cannot be used for data sets with irregular time spacing. According to Wright's modification, the extended form of this method for irregularly spaced data is:

$$\hat{y}_n = A_n V_n \tag{5}$$

where A_n is called the weighted average and α is a coefficient ranging from 0-1,

$$A_n = \sum_{i=-\infty}^{n} (1-\alpha)^{t_n - t_i} y_{t_i} \tag{6}$$

where V_n is called the normalizing factor, and

$$\frac{1}{V_n} = \sum_{i=-\infty}^{n} (1-\alpha)^{t_n - t_i} \tag{7}$$

The recursive form of Wright's modification is

$$\hat{y}_{t_n} = (1-V_n)\hat{y}_{t_{n-1}} + V_n y_{t_n} \tag{8}$$

$$V_n = \frac{V_{n-1}}{(b_n + V_{n-1})} \tag{9}$$

$$b_n = (1-\alpha)^{t_n - t_{n-1}} \tag{10}$$

Given the extended weighted average, the simple exponential smoothing method can be used to make forecasts using irregularly spaced data.

C. Wright's Modification of Holt's Method

Holt's method was devised to deal with time series trends. It uses observations to provide estimates of both level and slope. The recursive form of this algorithm is

$$L_t = \alpha y_t + (1-\alpha)(L_{t-1} + T_{t-1}) \tag{11}$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \tag{12}$$

where L_t and T_t are estimates of level and slope at time t. The forecast for time $t+k$ made at time t is

$$\hat{y}_t(t+k) = L_t + kT_t \tag{13}$$

One of the advantages of Holt's method is the flexibility it gives due to its two parameters. However, the form given above cannot be used for a dataset with irregular time spacing. The same modification was made to Holt's method:

$$y_{t_n}(t_n + k) = L_n + M_n k \tag{14}$$

$$L_n = (1-V_n)(L_{n-1} + (t_n - t_{n-1})M_{n-1}) + V_n y_{t_n} \tag{15}$$

$$M_n = (1-U_n)M_{n-1} + U_n(L_n - L_{n-1})/(t_n - t_{n-1}) \tag{16}$$

where

$$V_n = \frac{V_{n-1}}{(b_n + V_{n-1})} \tag{17}$$

$$b_n = (1-\alpha)^{t_n - t_{n-1}} \tag{18}$$

and

$$U_n = \frac{U_{n-1}}{(d_n + U_{n-1})} \tag{19}$$

$$d_n = (1-\beta)^{t_n - t_{n-1}} \tag{20}$$

V_n and U_n are the same as the normalizing factor V_n in the modified version of the simple exponential smoothing method. α and β are coefficients ranging from 0-1. The modified version of Holt's method still has two parameters and maintains its flexibility as a result.

Usefulness of the Acceleration Information

The ARIMA model was used to forecast the accelerations, so that the accelerations can be introduced for speed forecasting. Box and Jenkins introduced the use of ARIMA model and developed a methodology for data fitting [16].

D. The ARIMA Model

The ARIMA model of order (p, d, q) can be written using the back-shift operator notation:

$$\varphi(B)(1-B)^d(X_t - \mu) = \theta(B)\alpha_t \tag{21}$$

where

$$\varphi(B) = 1 - \varphi_1 B - \cdots - \varphi_p B^p \tag{22}$$

$$\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q \tag{23}$$

in which φ is an autoregressive parameter, θ is a moving average parameter, the back-shift operator is defined as follows:

$$BX_t = X_{t-1} \tag{24}$$

and α_t is white noise that can be defined as follows:

$$\alpha_t = (X_t - \mu) - \varphi_1(X_{t-1} - \mu) \tag{25}$$

While the ARIMA model is a general class of model, the autoregressive model and the moving average model can be viewed as special sub-classes of the ARIMA model.

Box and Jenkins method was used to develop an ARIMA model. According to Box and Jenkins method, the sample autocorrelation and partial sample autocorrelation diagrams should be examined first to determine the orders (p, d, q) of the ARIMA model [16]. The maximum likelihood method or the conditional least squares method can then be used to fit the model. However, these methods cannot deal with intermittent GPS data in which second-by-second observations are largely missing. Hence, in this study, the ARIMA model was used to forecast the acceleration after a transformation. The transformation transfers speed data measured by time into acceleration data measured by an index. The ARIMA model can therefore be used to forecast the acceleration.

Utilization of Acceleration Information in ARIMA models

Neural network was used to utilize the acceleration information to the speed prediction by aggregating the acceleration forecast from the ARIMA models and the speed forecast from the three existing parametric methods. To compare, a neural network without acceleration information and an extrapolation method using acceleration were also included.

E. Neural Network with Accelerations

The neural network can learn the relationship between the input and output by iteratively adjusting the weights in the network to minimize the difference between the observed input and observed output. The neural network used in this paper was a feedforward backpropagation network designed to have an input layer consisted of 4 neurons, a hidden layer with 12 neurons, and an output layer with one neuron. The number of neurons in the hidden layer was set to 12, which offers the best performance. The hyperbolic tangent sigmoid transfer function was used in the hidden layer, and the pure linear transfer function is used in the output layer. The Levenberg-Marquardt backpropagation learning algorithm was chosen to train the neural network. The maximum training epoch was set to be 1000.

F. Extrapolation method with Accelerations

The extrapolation method used the latest observed speed and the predicted acceleration to predict the next speed, which can be written as:

$$\hat{y}_t(t+k) = y_t + ka_t \tag{26}$$

where y_t is the speed observed at time t, a_t is the predicted acceleration.

G. Neural Network without Accelerations

This method intends to provide an aggregation of the result of naïve method, exponential smoothing method and Holt's method using neural network, which is compared with the method of neural network with acceleration. Therefore, the structure of the neural network in this section is identical to the structure of the neural network method with acceleration, except that there is no acceleration information input to the network, i.e., only three neurons in the input

layer were considered. The number of neurons in the hidden layer was also set to 12.

Measure of Effectiveness (MOE)

H. MOE

The mean absolute percentage error (MAPE) was used to measure the effectiveness of these methods. MAPE is a widely used standard to measure the effectiveness of forecasting methods by comparing the forecasting results to the observed values in terms of an absolute percentage. MAPE is defined as:

$$\text{MAPE} = \frac{1}{N} \sum_{n=1}^{N} \frac{\left| X_n - \bar{X}_n \right|}{X_n} \tag{27}$$

where X_n is the value observed at time n and $\bar{X}_n$ is the forecast for time n.

IV. RESULTS & DISCUSSION

A. ARIMA Model Order

The autocorrelation diagram and partial autocorrelation diagram of acceleration was examined to define the order of the ARIMA model.
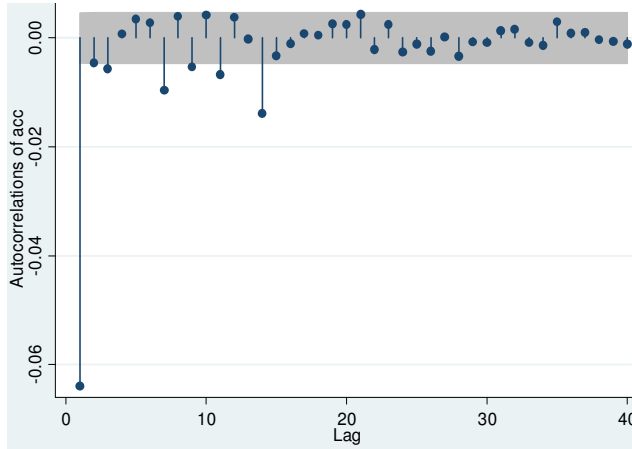

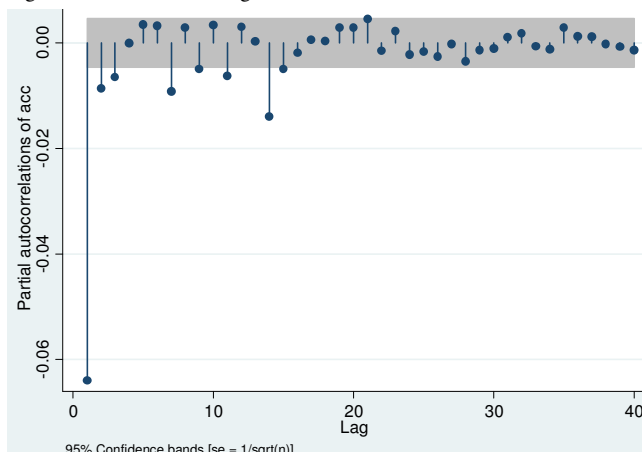Fig. 1. Autocorrelation diagram of acceleration

Fig. 2. Partial autocorrelation diagram of acceleration

STATA was used to calculate the autocorrelation diagram and the partial autocorrelation diagram for the sample acceleration. The diagrams are presented in Fig. 1 and Fig. 2.

In Fig. 1, for high lags, the sample autocorrelations are all very small and do not follow a very smooth pattern, so there is no need for differencing. The first sample autocorrelation is very large and the second one is slightly bigger than the confidence band, while the reminders are quite small, suggesting an MA(2) model for the series.

Similarly, in Fig. 2, the first three of the sample partial autocorrelations is moderately large, while the reminders are very small, suggesting an AR(3) model for the series.

Taking these two parameters into consideration, it was determined that it may also be useful to include ARMA (1, 1) and ARMA (2,1) in the analysis.

The Schwartz Bayesian criterion (SBC), often called the Bayesian information criterion (BIC), was used to measure which model best fit the sample series in Table I.

Table I
THE SBC OF 3 ARIMA MODELS WITH DIFFERENT ORDERS

ARIMA (p,0,q)	ARIMA (1,0,1)	ARIMA (3,0,0)	ARIMA (0,0,2)	ARIMA (2,0,1)
cons	-.00675	-.00675	-.00675	-.00675
ar1	.09435	-.06460		.33287
ar2		-.00909		.01612
ar3		-.00651		
ma1	-.15889		-.06464	-.39750
ma2			-.00526	
SBC	445595.6	445602.8	445596.5	445605.5

As in Table I, the ARIMA (1, 0, 1) has the smallest SBC. This ARIMA (1, 0, 1) was thus selected to make forecasts.

B. Comparison of Forecast Results

The six forecast methods were used to give a second-by-second forecast. The MAPE was calculated for time points at which there were observations. By comparing the MAPE of the results, the best performed method can be found, in which the MAPE indicates the extent to which the predicted value approaches the observed values. The comparison between different forecasting methods was shown in Table II.

The extrapolation method outperforms the three parametric methods, which indicates the essence of acceleration information in the estimation. The neural network without acceleration produces the worst results, which demonstrates that a larger number of parameters do not necessarily produce better results. A carefully designed scheme that utilizes the important characteristics of the problem is important for the development of a forecasting scheme. The neural network with acceleration information outperforms all of the other 5 methods, which utilizes the use of acceleration information and the aggregation of the three existing parametric methods. The comparison between neural network method with and without acceleration information shows that with nearly the same parameters, the acceleration information can greatly increase the forecast precision. It is envisaged that the acceleration information offers a useful indication of the phase transition in the

Table II

THE PERFORMANCE OF SIX DIFFERENT FORECASTING METHODS

Segment No.	Existing Parametric Methods			Proposed Methods		
	Naïve	Exponential Smoothing	Holt's Method	Neural Network without acceleration	Extrapolation method with acceleration	Neural Network with acceleration
Calibrated Parameters	None	alpha=0.004	Alpha=0.0009 Beta=0.0000004	Weights (not shown)	None	Weights (not shown)
1	3.60	3.24	2.50	3.31	3.34	1.78
2	3.42	3.39	3.10	3.70	3.19	2.07
3	2.07	2.09	2.24	2.41	1.74	1.47
4	2.09	2.09	2.17	2.49	1.76	2.35
5	1.38	1.36	1.48	1.61	1.18	0.85
6	2.47	2.35	2.28	2.76	2.42	1.74
7	2.82	2.70	2.55	3.04	2.36	2.13
8	5.80	6.15	6.48	7.30	6.68	9.62
9	5.86	6.13	5.96	7.28	4.89	4.29
10	3.49	4.22	5.77	5.20	3.07	2.79
11	3.24	3.15	3.30	4.19	2.40	2.66
12	2.48	2.55	2.96	3.51	2.06	2.15
13	4.73	4.92	5.61	7.76	3.86	4.39
14	1.14	1.09	1.11	1.28	1.18	0.81
15	1.26	1.52	1.70	1.54	1.28	1.21
16	1.12	1.13	1.37	1.59	1.09	1.09
17	1.20	1.06	0.95	1.24	1.08	0.88
18	0.46	0.44	0.49	0.73	0.51	0.47
19	1.69	1.66	1.86	2.12	1.47	1.15
20	5.01	5.13	5.14	5.64	4.42	2.82
Combined All	2.79	2.76	2.71	3.33	2.54	1.74

evolution of traffic conditions, and thus helps to more accurately predict the change in speed. Table II also shows that there is no obvious pattern of the prediction error across different road segments, thus indicating that there is no significant spatial correlation between these road segments.

V. CONCLUSION

The study reported here employed several methods to forecast traffic speeds using irregularly spaced GPS data from taxis, most of which were located on a 2-kilometer long section of Princess Margaret Road approaching the Cross Harbour Tunnel in Hong Kong. Three existing methods, the naïve method, the simple exponential method and Holt's method, were used to make forecasts from the irregularly spaced GPS data, while the ARIMA model was used to forecast acceleration. Three methods, neural network models with and without acceleration information and an extrapolation method, were proposed to aggregate the acceleration information into traffic speed forecasting and measure the usefulness of acceleration information. The MAPE was used to measure the forecasting performance of the methods. The results suggest that the neural network with the acceleration information outperforms all of the other methods. However, without the acceleration information, the

neural network method performs the worst. The introduction of acceleration information can improve the neural network forecasting results, because the resulting neural network forecasting utilizes the information of irregular time intervals and captures the phase transition in the speed forecasting. In future works, we will consider the model-based or fuzzy-neural approach to model the interaction between adjacent road segments (e.g., see [14], [17]).

REFERENCES

[1]   J. D. Croston, "FORECASTING AND STOCK CONTROL FOR INTERMITTENT DEMANDS," *Operational Research Quarterly,* vol. 23, pp. 289-&, 1972.

[2] R. H. Jones, "MAXIMUM-LIKELIHOOD FITTING OF ARMA MODELS TO TIME-SERIES WITH MISSING OBSERVATIONS," *Technometrics,* vol. 22, pp. 389-395, 1980.

[3] D. J. Wright, "FORECASTING DATA PUBLISHED AT IRREGULAR TIME INTERVALS USING AN EXTENSION OF HOLT METHOD," *Management Science,* vol. 32, pp. 499-510, Apr 1986.

[4] T. Cipra*, et al.*, "HOLT-WINTERS METHOD WITH MISSING OBSERVATIONS," *Management Science,* vol. 41, pp. 174-178, Jan 1995.

[5] J. L. Carmo and A. J. Rodrigues, "Adaptive forecasting of irregular demand processes," *Engineering Applications of Artificial Intelligence,* vol. 17, pp. 137-143, Mar 2004.

[6] T. R. Willemain*, et al.*, "A new approach to forecasting intermittent demand for service parts inventories," *International Journal of Forecasting,* vol. 20, pp. 375-387, Jul-Sep 2004.

[7] W. F. Velicer and S. M. Colby, "A comparison of missing-data procedures for ARIMA time-series analysis," *Educational and Psychological Measurement,* vol. 65, pp. 596-615, Aug 2005.

[8] T. Cipra, "Exponential smoothing for irregular data," *Applications of Mathematics,* vol. 51, pp. 597-604, 2006.

[9] N. Altay*, et al.*, "Adapting Wright's modification of Holt's method to forecasting intermittent demand," *International Journal of Production Economics,* vol. 111, pp. 389-408, Feb 2008.

[10] T. Cipra and T. Hanzak, "Exponential smoothing for irregular time series," *Kybernetika,* vol. 44, pp. 385-399, 2008.

[11] B. L. Smith and M. J. Demetsky, "Traffic flow forecasting: Comparison of modeling approaches," *Journal of Transportation Engineering-Asce,* vol. 123, pp. 261-266, Jul-Aug 1997.

[12] B. L. Smith*, et al.*, "Comparison of parametric and nonparametric models for traffic flow forecasting," *Transportation Research Part C-Emerging Technologies,* vol. 10, pp. 303-321, Aug 2002.

[13] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results," *Journal of Transportation Engineering-Asce,* vol. 129, pp. 664-672, Nov-Dec 2003.

[14] W. Y. Szeto*, et al.*, "Multivariate Traffic Forecasting Technique Using Cell Transmission Model and SARIMA Model," *Journal of Transportation Engineering-Asce,* vol. 135, pp. 658-667, Sep 2009.

[15] M. C. Tan*, et al.*, "An Aggregation Approach to Short-Term Traffic Flow Prediction," *Ieee Transactions on Intelligent Transportation Systems,* vol. 10, pp. 60-69, Mar 2009.

[16] G. E. P. Box, *Time series analysis; forecasting and control [by] George E. P. Box and Gwilym M. Jenkins*. San Francisco: Holden-Day, 1970.

[17] H. B. Yin*, et al.*, "Urban traffic flow prediction using a fuzzy-neural approach," *Transportation Research Part C-Emerging Technologies,* vol. 10, pp. 85-98, Apr 2002.